

## Predictive modelling method application to computer-aided medical diagnosis

**Publication number:** CN1477581 (A)

**Publication date:** 2004-02-25

**Inventor(s):** ZHOU ZHIHUA [CN]

**Applicant(s):** UNIV NANJING [CN]

**Classification:**

- **international:** **G06F19/00; G06F19/00;** (IPC1-7): G06F19/00

- **European:**

**Application number:** CN20031032141 20030701

**Priority number(s):** CN20031032141 20030701

**Also published as:**

 CN1234092 (C)

### Abstract of **CN 1477581 (A)**

The present invention discloses a prediction model-building method for computer-aided medical diagnosis. It utilizes the medical symptom detection equipment to obtain the symptom of object to be diagnosed to form symptomatic vector, then utilizes the prediction model to make treatment so as to can obtain prediction result. It also provides its concrete steps for implementing said invented prediction model-building method.

---

Data supplied from the **esp@cenet** database — Worldwide



## [12] 发明专利申请公开说明书

//G06F159:00, A61B5/00,  
G06N3/02

[21] 申请号 03132141.0

[43] 公开日 2004年2月25日

[11] 公开号 CN 1477581A

[22] 申请日 2003.7.1 [21] 申请号 03132141.0

[71] 申请人 南京大学

地址 210093 江苏省南京市汉口路22号

[72] 发明人 周志华

[74] 专利代理机构 南京苏高专利事务所

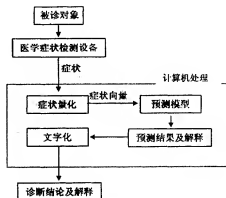
代理人 柏尚春

权利要求书1页 说明书3页 附图3页

[54] 发明名称 一种适用于计算机辅助医疗诊断的  
预测建模方法

## [57] 摘要

本发明公开了一种适用于计算机辅助医疗诊断的预测建模方法,包括通过医学症状检测设备获取待诊对象的症状形成症状向量,经预测模型处理,即可得到预测结果,该方法包括以下步骤:(1)若预测模型未训练好,则执行步骤(2),否则转到步骤(6);(2)利用历史病例产生初始训练数据集;(3)利用初始训练数据集训练出一个神经网络集成;(4)利用神经网络集成对初始训练数据集进行处理以产生规则训练数据集;(5)利用规则学习技术从规则训练数据集中产生规则模型;(6)利用规则模型进行预测并给出结果及解释。本发明的优点是:为计算机辅助医疗诊断装置提供了一种高精度、高可理解性的预测建模方法。



1、一种适用于计算机辅助医疗诊断的预测建模方法，包括通过医学症状检测设备获取待诊对象的症状，然后将症状进行量化得到症状向量 $[t_1, t_2, \dots, t_n]$ ，其中 $t_n$ 表示第 $n$ 个症状值，症状向量交给预测模型处理，即可得到预测结果及解释的数字化表示形式，其特征是该方法包括以下步骤：

- (1) 若预测模型未训练好，则执行步骤(2)，否则转到步骤(6)；
- (2) 利用历史病例产生初始训练数据集；
- (3) 利用初始训练数据集训练出一个神经网络集成；
- (4) 利用神经网络集成对初始训练数据集进行处理以产生规则训练数据集；
- (5) 利用规则学习技术从规则训练数据集中产生规则模型；
- (6) 利用规则模型进行预测并给出结果及解释；
- (7) 结束。

2、根据权利要求1所述的适用于计算机辅助医疗诊断的预测建模方法，其特征是：在(4)中，利用神经网络集成产生用于建立规则模型的规则训练数据集 $L_i$ 的步骤是：

- (1) 将 $L_i$ 置为空集；
- (2) 从初始训练数据集 $L_0$ 中获取一个症状向量及其类别；
- (3) 为每个类别分别设置一个计数器，用来记录神经网络给出的同类别预测结果的数目；
- (4) 将所有计数器清零；
- (5) 将控制参数 $k$ 置为1， $k$ 是一个大于等于1但小于等于神经网络集成中神经网络的个数 $N$ ；
- (6) 取得神经网络集成中第 $k$ 个神经网络对待诊症状向量给出的预测结果 $F_k$ ；
- (7) 将 $F_k$ 所对应的类别的计数器加1；
- (8) 将 $k$ 加1；
- (9) 判断 $k$ 是否小于等于神经网络集成中神经网络的个数 $N$ ，如果是则表明还有其他神经网络尚未考察，转到步骤(6)；否则执行步骤(10)；
- (10) 对所有计数器中的值进行比较，找出值最大的计数器，并将其对应的类别作为当前症状向量的新类别；如果有多个计数器中的值均为最大值，则以这些计数器对应的类别中出现机会最大的疾病种类作为当前症状向量的新类别；
- (11) 将当前症状向量及其新类别加入 $L_i$ ；
- (12) 判断 $L_0$ 中是否还有未考察的症状向量，如果有则转到步骤(2)；否则进入步骤(13)；
- (13) 结束。

## 一种适用于计算机辅助医疗诊断的预测建模方法

### 一、技术领域

本发明涉及一种计算机辅助医疗诊断装置，特别涉及一种利用神经网络集成技术和规则学习技术的高精度、高可理解性预测建模方法。

### 二、背景技术

随着计算机技术的发展，计算机辅助医疗诊断装置由于不受疲劳、情绪等因素的影响，已成为重要的辅助诊断手段。计算机辅助医疗诊断装置通常是利用一些预测建模方法对历史病例进行分析，从而建立预测模型，然后再用该预测模型来对新病例进行诊断，其结果提交给医学专家进行进一步的分析确诊，从而在一定程度上减轻医学专家的工作负担。因此，预测建模方法是计算机辅助医疗诊断装置的关键。一方面，由于医疗诊断务求精确，因此适用的预测建模方法必须具有很高的精度；另一方面，由于医疗诊断事关被诊者的身体健康和生命安全，因此适用的预测建模方法必须具有很高的可理解性，即在作出诊断结论之后还需要能提供对诊断的解释，这不仅是被诊者及其家属的需要，还是医学专家检查诊断过程的需要。然而，现有技术如神经网络等虽然具有高精度，但不具有高可理解性；而规则学习等虽然具有高可理解性，但却不具有高精度，这就对计算机辅助医疗诊断装置的性能造成了不利影响。

### 三、发明内容

本发明的目的是针对现有技术难以产生适用于计算机辅助医疗诊断装置的高精度、高可理解性预测模型的问题，提供一种高精度、高可理解性的预测建模方法，以辅助提高计算机辅助医疗诊断装置的性能。

为实现本发明所述目的，本发明提供一种利用机器学习中的神经网络集成技术和规则学习技术进行预测建模的方法，该方法包括以下步骤：（1）若预测模型未训练好，则执行步骤2，否则转到步骤6；（2）利用历史病例产生初始训练数据集；（3）利用初始训练数据集训练出一个神经网络集成；（4）利用神经网络集成对初始训练数据集进行处理以产生规则训练数据集；（5）利用规则学习技术从规则训练数据集中产生规则模型；（6）利用规则模型进行预测并给出结果及解释；（7）结束。

本发明的优点是：为计算机辅助医疗诊断装置提供了一种高精度、高可理解性的预测建模方法，以辅助提高计算机辅助医疗诊断装置的性能。

下面将结合附图对最佳实施例进行详细说明。

### 四、附图说明

图1是计算机辅助医疗诊断装置的工作流程图。

图2是本发明方法的流程图。

图3是用神经网络集成产生规则训练数据集的流程图。

## 五、具体实施方式

如图1所示,计算机辅助医疗诊断装置利用医学症状检测设备例如体温、血压测量设备等获取待诊对象的症状例如体温、血压等,然后将症状进行量化以得到症状向量,例如 $[t_1, t_2, \dots, t_n]$ ,其中 $t_1$ 表示第一个症状值, $t_2$ 表示第二个症状值,依此类推。症状向量交给预测模型处理,即可得到预测结果及解释的数字化表示形式,经过文字化处理后,就产生了最后提交给用户的诊断结论及解释。

本发明的方法如图2所示。步骤10是初始动作。步骤11判断预测模型是否已经训练好,若已训练好则可处理诊断任务,执行步骤16;否则需进行训练,执行步骤12。步骤12利用历史病例产生初始训练数据集,为叙述方便,称初始训练数据集为 $L_0$ 。 $L_0$ 中包含了每一历史病例所对应的症状向量及其类别,即诊断出的具体疾病类别(“没有疾病”也作为一种类别)。步骤13利用统计学中常用的可重复取样技术从 $L_0$ 中产生 $N$ 个数据集,并用这 $N$ 个数据集中的一个训练出一个神经网络,这些神经网络就组成了神经网络集成。 $N$ 是一个用户预设的整数值例如9,它确定了神经网络集成所包含的神经网络个数。这里使用的神经网络可以是任何类型的神经网络,只要可以执行预测任务即可,例如可以使用神经网络教科书中介绍的多层前馈BP网络。步骤14利用神经网络集成产生用于建立规则模型的规则训练数据集 $L_1$ ,该步骤将在后面的部分结合图3进行具体介绍。

图2的步骤15利用 $L_1$ 训练出规则模型。规则模型是一个由很多条IF-Then或类似形式的规则组成的预测模型,它由某种规则学习方法从某个训练数据集(这里就是 $L_1$ )中训练出来。这里可以使用任何类型的规则学习方法,只要其产生的模型可以执行预测任务即可,例如可以使用机器学习教科书中介绍的RIPPER、C4.5 Rule等。步骤16接收待诊断的症状向量。步骤17将症状向量提交给训练好的规则模型进行预测。步骤18给出规则模型产生的预测结果及预测过程中使用的规则,这些规则就组成了对该预测结果的解释。步骤19是结束状态。

由于本发明的方法建立的预测模型是规则模型,因此其具有高理解性;又由于该方法利用了具有高精度的神经网络集成来产生建立规则模型的训练数据集,这可以视为对初始数据集进行了去噪、增强等良性处理,因此建立的规则模型也具有高精度。

图3详细说明了图2的步骤14,其作用是利用神经网络集成来产生用于建立规则模型的规则训练数据集 $L_1$ 。图3的步骤140是起始状态。步骤141将 $L_1$ 置为空集。步骤142从图2的步骤12产生的初始训练数据集 $L_0$ 中获取一个症状向量及其类别。步骤143为每个类别分别设置一个计数器,这些计数器用来记录有多少个神经网络给出的预测结果是该类别,这里的各类别分别对应了诊断出的具体疾病类别(“没有疾病”也作为一种类别)。

步骤 144 将所有计数器清零。步骤 145 将控制参数  $k$  置为 1,  $k$  是一个大于等于 1 但小于等于图 2 中步骤 13 的  $N$  的一个整数值, 它用来指示当前考察的神经网络的序号。步骤 146 取得神经网络集成中第  $k$  个神经网络对待诊症状向量给出的预测结果, 为叙述方便, 称该结果为  $F_k$ 。步骤 147 将  $F_k$  所对应的类别的计数器加一。步骤 148 将  $k$  加一。步骤 149 判断  $k$  是否小于等于神经网络集成中神经网络的个数, 即图 2 中步骤 13 的  $N$ , 如果是则表明还有其他神经网络尚未考察, 转到步骤 146; 否则就执行步骤 150。

图 3 的步骤 150 对所有计数器中的值进行比较, 找出值最大的计数器, 并将其对应的类别作为当前症状向量的新类别; 如果有多个计数器中的值均为最大值, 则以这些计数器对应的类别中出现机会最大的疾病种类作为当前症状向量的新类别。步骤 151 将当前症状向量及其新类别加入  $L_i$ 。步骤 152 判断  $L_0$  中是否还有未考察的症状向量, 如果有则转到步骤 142; 否则就进入步骤 153, 即图 3 的结束状态。

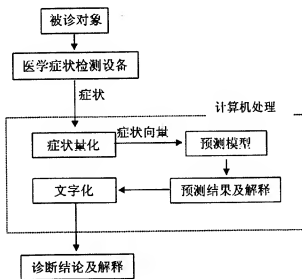


图 1

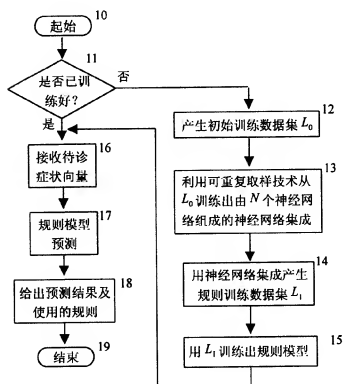


图 2



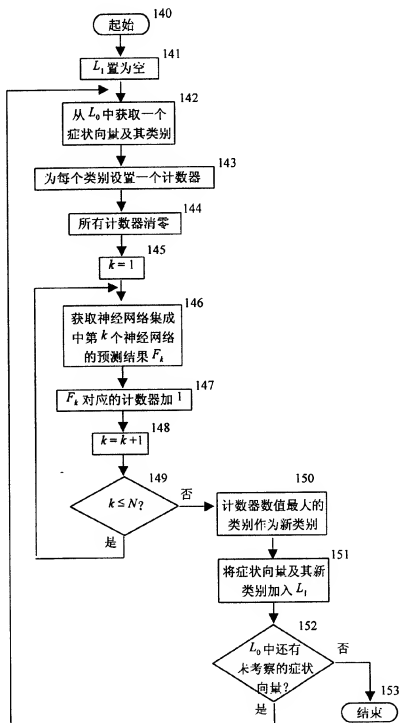


图 3